Weliky, M., and Katz, L.C. (1999). Science *285*, 599–604.
Wiesel, T.N., and Hubel, D.H. (1965). J. Neurophysiol. *28*, 1060–1072.
Zheng, W., and Knudsen, E.I. (1999). Science *284*, 962–965.

DOI 10.1016/j.neuron.2005.10.007

Attractor Neural Networks and Spatial Maps in Hippocampus

Attractor neural network theory has been proposed as a theory for long-term memory. Recent studies of hippocampal place cells, including a study by Leutgeb et al. in this issue of *Neuron*, address the potential role of attractor dynamics in the formation of hippocampal representations of spatial maps.

Attractor networks have been proposed as a mechanism by which the brain is able to encode and store memories and representations of the external world. The basic tenants of attractor network theory as it applies to the encoding of long-term memories are as follows (Hopfield, 1982).

(1) Long-term memories reside in specialized recurrent neural networks as specific patterns of neuronal activity; (2) the memory states are imprinted in the network with long-lasting modifications of recurrent synaptic connections of the Hebbian type; (3) as a result of synaptic modifications, the memory states can be retrieved through input of partial cues and persist due to recurrent self-excitation without a further need for stabilizing inputs, i.e., turn into attractor states of the network. The appealing feature of attractor networks as a model idea is that this theory provides a unified description of several crucial aspects of memory, such as encoding, storage, retrieval, long-term and working memory, etc. Although attractor networks have been largely the product of theoretical modeling, experimental studies support these ideas, mainly in the form of persistent delay activities of neocortical neurons observed during memory experiments in primates (Fuster and Alexander 1971, Miyashita 1988, Miller 1996, Williams and Goldman-Rakic 1995). Indeed, delayed activities are observed when the sensory inputs that gave rise to them are withdrawn and presumably come about due to recurrent self-excitation resulting from synaptic modifications in the learning stage, as proposed in attractor neural network models (Amit and Mongillo 2003).

The hippocampus has been implicated as the intermediate storage place for episodic memories, based largely on observing the effects of hippocampal lesions in human patients (see Murray, 2000, for the critical analysis of these initial observations). In particular, it has been suggested that memories are stored in area CA3 of the hippocampus (Treves and Rolls, 1994). CA3 is a region characterized by the heavy recurrent connectivity, which is a crucial ingredient of the attractor networks. On the other hand, a large body of work shows that hippocampal pyramidal neurons in rodents are selectively activated at specific locations in an environment during exploration (O'Keefe and Dostrovsky, 1971), and it is now possible to record many hippocampal neurons simultaneously in the awake behaving animals, facilitating the search for attractor activity. Place cells are cells in the hippocampus that are activated when an animal passes through a specific location (the "place field") in its environment. Different groups can be active in different environments, or in some cases, the same cell can fire in different places fields. When animals are exposed to different environments, the representation of place fields change, a process known as "remapping." (Lever et al., 2002).

One important question that the previous remapping work raised is whether hippocampal representations in different environments represent different attractor states. The search for attractor states in the hippocampus has met with a number of challenges. Since hippocampus is a multimodal integration area and hippocampal place cells are driven by a variety of sensory inputs and intrinsically generated path-integration signals, one considerable hurdle is to design a controlled situation where the hippocampus is disconnected from all external influences. Nonetheless, a large amount of indirect evidence has accumulated to suggest that attractor-like networks may underlie the spatial maps in the hippocampus (Tsodyks, 1999).

Recently, strong support for the attractor hypothesis came from a study by Wills et al. that recorded place field activity in behaving rats exposed to different environments that were gradual transformations between two basic shapes—a circle and a square (Wills et al., 2005). Rats were presented with two environments that differed in shape, color, and texture (a wooden circle and a square "morph box"), and the responses of place cells were rapidly remapped to differentiate between the two environments. To test how the place field representations changed when animals were in intermediate environments, animals were tested in the "morph box," the shape of which could be changed to make the environment more circular or more square-like. Wills et al. reported that when place fields are recorded in these morphed environments, most cells fired in a pattern that was either circle-like or square-like. To explain how this study relates to attractor networks, imagine that spatial maps in these two environments represent two distinct attractor states in the hippocampal network. Whatever the extrahippocampal inputs that give rise to these states are, one should expect that if they are gradually transformed (morphed) from one to another along a certain transformation trajectory, a point should be reached where basins of attraction of these two attractors meet. At this point, small changes in the input should produce dramatic changes in the activity state that is being pushed to either of the two attractors due to intrinsic network dynamics. In other words, if there are two distinct attractors, then the model would predict a sharp, coherent transition of the network activity as one moves between the morphed environments. Indeed, this is what was reported in the Wills et al. study—many of the hippocampal neurons changed their place correlates sharply and coherently at a certain position along the morph sequence (Wills et al., 2005).

The study in this issue of *Neuron* by Leutgeb et al. (2005) took a similar experimental approach, although

with some twists, and arrived at somewhat different conclusions. Like the Wills et al. study, rats were first trained to search for food in two environments—a circle and a square. In the case of the Wills et al. study, the initial training environments were quite distinct in terms of shape, color, and texture. The Leutgeb study trained animals in square and circle versions of the same arena (so presumably the only variable that changed between the environments was shape) until the two environments evoked very different CA3 network activation patterns. The authors then recorded ensemble activity in select pyramidal neurons in CA1 and CA3 while the original arena was morphed to the other (square to circle and vice versa) through a series of gradual intermediate shapes. As opposed to the results of the Wills et al. (2005) study, where the transition between network states was sharp and coherent, the authors reported gradual and incoherent changes in the mapping for different neurons along the morphing sequence between the square and the circle.

On the surface, the results from these two studies appear quite incongruous. However, attractors may also underlie these results as well. The results suggest that hippocampal network assemblies can exist on a continuum and may argue against the existence of a discrete global attactor. The results could be explained by local attractor states which act in a more continuous fashion. The most critical observation of their study is that when CA3 spatial maps for two environments are compared before and during the morphing experiments (that is before and after learning), drastically different results are obtained. After the rats were familiarized with the circular and square arenas, CA3 maps for them were almost completely decorrelated, i.e., different groups of hippocampal cells were active in these two arenas. However, when the rats are able explore the gradually transformed arenas, the corresponding groups of neurons change gradually and the overall difference accumulated between the square and circular arenas is much smaller, indicating that highly overlapping groups of neurons are active in these two environments. The observed change is always in the direction of the initial shape, i.e., when the rat moves from the circle to square, the later arena is now represented by the population that is intermediate between the two original groups mapping the circle and the square. This hysteresis type of behavior means that the representation of the current environment is sensitive not only to its geometrical properties but to the previous experience of the animal as well. More precisely, when moving from one environment into the other, the representation of the second environment is biased toward the activity patterns that represent the first one. One could reasonable argue that exploring the first environment led to the strengthening of the corresponding attractor state, which consequently shifted the representation of the second environment. Interestingly, the hysteresis effects weaken with repeated explorations of morphed environments, that is, the overlap between the representations of circle and square arenas diminishes with time, possibly due to the saturation of the synaptic plasticity in familiar environments. These surprising observations point to highly dynamic and flexible neural representations of environment in hippocampus, sensitive to the general behavioral context and adjusting to changing exploration history. Therefore, the seeming contradiction between the results obtained in Wills et al. (2005) and Leutgeb et al. (2005) could be due to precisely this flexibility: while in Wills et al. (2005) the rats explored the morphed environments in a scrambled order, in Leutgeb et al. (2005) sequential morphing was employed. These two exploration histories could lead to very different dynamics of spatial representations of the morphed environments. It appears that recent exciting developments uncovered a wealth of information for contemplating attractor neural networks in hippocampus and provided new directions for future research. Is it certain that distinct maps in the Wills et al. study represent global attractors in hippocampus? An alternative explanation would be that they arise due to the intricate modifications in the extrahippocampal inputs during the training. What are the underlying mechanisms that lead to the hysteresis behavior in the Leutgeb et al. study? The most plausible mechanism at this stage appears to be intrinsic synaptic plasticity in the CA3 area, but novel experimental techniques may have to be developed before this hypothesis can be tested directly. What is clear is that an efficient dialog between experimental and theoretical studies will play an increasingly important role in understanding the nature of hippocampal place representation.

Misha Tsodyks
Department of Neurobiology
Weizmann Institute of Science
Rehovot 76100
Israel

Selected Reading

Amit, D.J., and Mongillo, G. (2003). Cereb. Cortex *13*, 1139–1150.

Fuster, J.M., and Alexander, G.E. (1971). Science *173*, 652–654.

Hopfield, J.J. (1982). Proc. Natl. Acad. Sci. USA *79*, 2554–2558.

Leutgeb, J.K., Leutgeb, S., Treves, A., Meyer, R., Barnes, C.A., McNaughton, B.L., Moser, M.-B., and Moser, E.I. (2005). Neuron *48*, this issue, 345–358.

Lever, C., Wills, T., Cacucci, F., Burgess, N., and O'Keefe, J. (2002). Nature *416*, 90–94.

Miller, E.K. (1996). J. Neurosci. *16*, 5154–5167.

Miyashita, Y. (1988). Nature *335*, 817–820.

Murray, E.A. (2000). In The New Cognitive Neurosciences, M.S. Gazzaniga, ed. (Cambridge, MA: The MIT Press).

O'Keefe, J., and Dostrovsky, J. (1971). Brain Res. *34*, 171–175.

Treves, A., and Rolls, E.T. (1994). Hippocampus *4*, 374–391.

Tsodyks, M. (1999). Hippocampus *9*, 481–489.

Williams, V.W., and Goldman-Rakic, P.S. (1995). Nature *376*, 572–575.

Wills, T.J., Lever, C., Cacucci, F., Burgess, N., and O'Keefe, J. (2005). Science *308*, 873–876.

DOI 10.1016/j.neuron.2005.10.006